基于 ChatGPT 的用户图书评分偏好预测研究

陈燕方 [1]，李志宇 [2]
（1. 中国人民大学图书馆，北京，100872；2. 北京科学智能研究院，北京，100084）

摘要：[目的/意义] 随着以 ChatGPT 为代表大语言模型技术的不断发展与变革，使得许多领域的经典场景都重新焕发出新的机会。同时，越来越多的学者开始关注如何将大语言模型的智能化能力与技术应用到现有的场景，并分析这些技术带来的挑战和机遇。[方法/过程] 本文以 ChatGPT 为建模对象，首次将大语言模型技术引入用户图书评分偏好预测这一图情领域的典型应用场景，并落地实践。通过构建基于 ChatGPT 的用户图书评分预测模型（CUBR, ChatGPT-based model for User Book Rating Prediction），来探索大语言模型技术在图书推荐领域实践和落地的可行性。同时，本文基于图书评分任务的不同评估方案与现有经典推荐模型进行对比，探讨并给出了 CUBR 在用户图书评分预测场景的优势与劣势，并分析了后续大语言模型在图书推荐其他场景可能的研究机会点。[结果/结论] 本文实验研究表明：（1）CUBR 模型在现有用户图书评分偏好预测任务上能够取得不错的推荐效果，特别是单样本 (One-shot) 这类待推荐目标信息较少的情况下，其表现接近或超过当前经典推荐算法，且泛化能力较强，较适用于冷启动推荐场景。（2）随着单个用户提示样本内容的增加（如从 One-shot 到 Ten-shot），CUBR 的预估效果会有显著的提升，说明 CUBR 具备不错的实时上下文学习能力。[局限] 本文研究场景仅限于用户图书评分偏好理解与推荐，未来将尝试在更多的图情场景应用和改造现有大语言模型技术，并获得更好的实践效果。
关键词：ChatGPT；大语言模型；图书评分；生成式会话

A ChatGPT-based Model for User Book Rating Prediction

YanFang Chen[1]，Zhiyu Li[2]
(1. Renmin Unversity of China, Beijing, 100872, China; 2. AI for Science Institute, Beijing, 100084, China)

Abstract: [Purpose/significance] With the continuous development and change of Large Language Models (LLMs) represented by ChatGPT, classical scenarios in many fields have been given new opportunities. At the same time, more and more researchers begin to focus on how to apply the intelligentness and technology of LLMs to existing scenarios, and analyze the challenges and opportunities brought by these technologies. [Method/process] This is the first time that LLM technology has been introduced into user book rating prediction, which is a typical application scenario in library and information science. We explored the feasibility of using LLM technology in user book rating by building a CUBR (ChatGPT-based model for User Book Rating Prediction) model based on ChatGPT. At the same time, this paper compares different evaluation schemes based on book rating task with existing classical recommendation models, discusses and gives the advantages and disadvantages of CUBR in predicting scenarios of user book scoring, and analyses the possible application opportunities of subsequent LLMs in other scenarios of book recommendation. [Result/conclusion] The experimental research in this paper shows that: (1) CUBR model can achieve good recommendation results on existing user book rating prediction tasks, especially when the target information to be recommended is less, such as one-shot, which performs close to or exceeds the current classical recommendation algorithm, and has strong generalization ability, which is suitable for cold-start recommendation. (2) With the increase of sample content prompted by a single user (e.g. from One-shot to Ten-shot), the predictive effect of CUBR will be significantly improved, indicating that CUBR has good real-time in-context learning ability. [Limitations] The scenarios studied in this paper are limited to the understanding and recommendation of users' book scoring preferences. In the future, we will try to apply and transform the existing large language model technology in more library and information science scenarios, and achieve better landing effects.
Keywords: ChatGPT; Large Language Model; Book Rating; Generative Response;

* 本文为中国人民大学科学研究基金（中央高校基本科研业务费专项资金资助）项目（编号：23XNQT24）研究成果之一。

1 引言

近年来，随着自然语言处理技术（NLP，Natural Language Processing）的飞速发展，无论是从模型参数的规模，还是从训练数据的丰富程度来看，NLP 技术都发生着日新月异的变化。2022 年 12 月初，OpenAI 的发布了基于 GPT-3.5 系列大语言模型[1]（大语言模型, Large Language Models）构建并微调后的聊天对话机器人 ChatGPT[2]（Chat Generative Pre-trained Transformer）。该模型不仅能够针对多轮次对话，进行高效且精准的交互式回答，同时还能够进行包括辅助代码编写、文档摘要、小说续写等各类自然语言处理任务。该模型一经推出，随机在产业界和学术界引起了热烈的讨论。

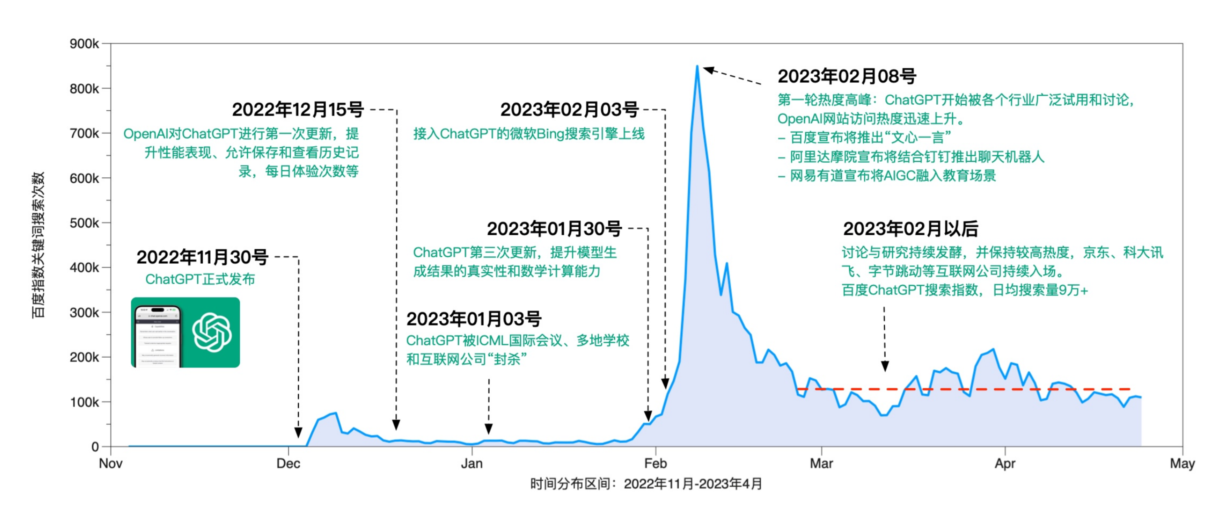


图 1 ChatGPT 百度搜索指数趋势图与典型的重要事件

如图1所示，为 ChatGPT 从 2022 年 12 月初发布至今的百度搜索指数变化趋势图与典型的重要事件标注。从图1中可以看到，在发布的初期阶段（2022 年 11 月-2023 年 2 月），由于模型的效果和用户界面的不完善，ChatGPT 的整体热度保持在相对低位的状态。2 月开始，随着 OpenAI 对模型的迭代，以及若干重要事件的报道，如 ChatGPT 通过谷歌工程师面试[3]、各个大型互联网公司的广泛参与，以 ChatGPT 所包含的技术底座：大语言模型模型，也受到了学术界的极大关注[1]。

当前流行的 LLM 模型版本主要包括 GPT3/4 系列模型[4]、LLaMA 模型[5]，以及国内清华大学推出的 GLM130B 模型[6] 等等。其中，依赖于微软以及 OpenAI 的大力推广与宣传，且支持 API 的进行模型调用，构建于 GPT3/3.5/4 系列模型之上的 ChatGPT 应用被越来越多的研究者和厂商接入和使用在各类现实场景。如智能客服[7]、交互翻译[8]、私人助手[9] 等等。

特别的，在图情领域，围绕 ChatGPT 类模型的理论研究也日趋增长。如 ChatGPT 类模型在图情领域的技术伦理与风险研究[10-11] 与应用场景研究[12-15] 等等。但上述的研究主要围绕从理论探讨到应用场景的分析，并未将 ChatGPT 落地到实际的应用上进行测试，例如是否能考虑基于 ChatGPT 类模型构建推荐模型来解决原有图情领域的各类推荐问题？ChatGPT 类模型应用到对应的推荐场景之后相比传统的推荐模型效果怎么样？都是非常值得探讨的。

本文的创新点主要体现在以下方面：（1）本文针对图情领域的典型任务——用户图书评分偏好预测任务，应用大语言类模型（如 ChatGPT）来构建相应的预测模型，以此探索 ChatGPT 类模型在图情领域落地和应用的可能性。（2）基于用户图书评分偏好任务，设计了对应预测建模所需的提示工程范例，给类似场景的相关落地研究带来一定的启发。（3）基于单样本和少样本建模，并在 GoodBook 数据集上进行实验，通过不同的实验指标论证了基于 ChatGPT 类大语言模型应用于用户图书评分偏好预估场景的可行性。

2 相关研究

随着馆藏资源的不断发展，无论是图书数量、种类，还是与读者的交互情况，都呈现出迅速增加的态势，因此图书馆的服务形式也不断朝着智慧化的方向发展，而推荐模型正是解决这一信息过载问题的重要手段与方法之一。其中，面向读者的图书资源的个性化推荐系统也多是借助机器学习推荐模型实现，围绕这一方向的部分研究包括：

基于读者（User）-图书（item）的协同过滤推荐模型：如余以胜等[16]，通过在基于物品协同过滤模型中引入偏置，结合偏置本身的含义和相似图书对预测评分的贡献，来改善图书推荐系统中的可解释性与准确性。而杨辰等[17] 则除了考虑图书内容本身的相似性之外，进一步引入用户的社交关系层面的相似性，并基于启发式的非监督方法来融合相似性度量，以此优化推荐效果。

基于读者-图书交互行为序列推荐模型：从单个用户角度上看，用户与图书的系列交互行为随着时间的变化能够形成对应的图书序列，该类模型主要解决的是下一（多）次时间可能交互对象的推荐问题[18]。例如，王代琳等[19] 提出一种基于图书目录注意力机制的个性化推荐模型，借助用户评分和注意力机制，对用户的历史浏览交

互行为进行建模，基于 BiLSTM 融入读者的兴趣偏好，以此提高推荐的准确性，但不足之处在于该模型表现强依赖于稠密的读者行为矩阵，在稀疏场景下效果受限。

基于读者-图书网络的图神经网络推荐模型。受益于图模型的特征表达与高阶抽取能力，图神经网络已被广泛的应用于推荐系统的各个方向。如陈帜等[20]基于图卷积神经网络对读者-图书二部图构建的交互历史进行建模，捕捉节点之间的高阶连通性来更好的建模读者的领域偏好信息，提高推荐效果。

3 基于 ChatGPT 类的用户图书评分偏好预估模型

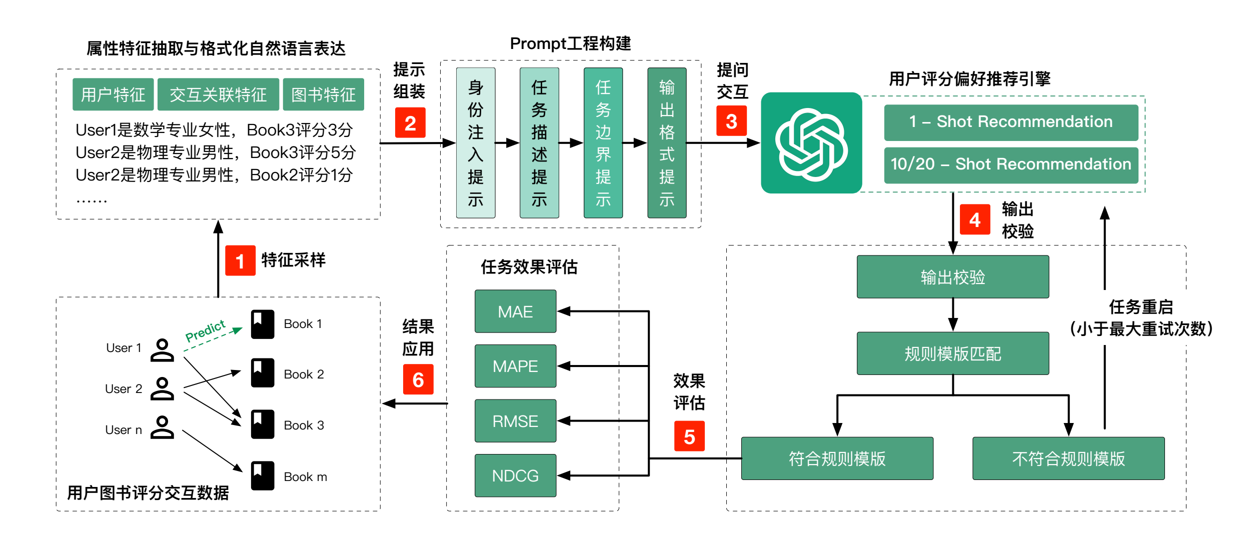


图 2 用户图书评分偏好预估模型框架

3.1 模型概述

本文提出了一种基于 ChatGPT 类的大语言模型的用户图书评分偏好预估模型，该模型通过将现有的大语言模型与用户评分偏好预估任务相结合，构造合适的 Prompt 策略，并结合数据校验、回溯与重试方法，最终探索 LLM 在用户图书评分偏好预估场景应用的可能性。模型整体分为四个模块：(1) 任务形式化定义。(2) 任务提示工程设计（Prompt Engineering）。(3) 模型交互与响应解析与校验。(4) 任务指标评估。

3.2 任务形式化定义

用户评分偏好预测是根据用户与图书的历史交互或评分行为，对用户未来时刻可能与其他图书产生交互的偏好进行预估。该任务在图书推荐领域应用场景非常广泛，例如，面向电商销售场景的用户图书偏好预估，面向图书馆读者图书借阅、点击、浏览兴趣偏好预估推荐等。该任务通常以读者与图书的历史交互（点击、浏览、借阅、收藏、评论、打分等）作为特征与数据来源，结合用户基础属性与图书属性等，利用多种机器学习模型来构建精准的推荐。本文中，具体任务定义如下：

用户单样本推荐建模：给定用户 u_i 的历史图书行为样本序列（如评分序列）：$H_{u_i} = \{b_1, b_2, ..., b_n\}$，仅给模型提供单个训练样本作为提示或训练集，要求模型对行为序列中的剩余全部样本进行偏好打分，最终评估模型打分结果与原始样本结果的一致性。

用户少样本推荐建模：给定用户 u_i 的历史图书行为序列（如评分序列）：$H_{u_i} = \{b_1, b_2, ..., b_n\}$，从中选出一定比例的数据作为训练集（本文中将分别选取 10 个、20 个提示样例作为提示集）(或 Prompt 提示集)，要求模型对剩余的序列进行偏好打分，最终评估模型打分结果与原始样本结果的一致性。

3.3 任务提示工程设计

由于 ChatGPT 类大语言模型是一种典型的生成式模型，其生成内容的质量好坏通常取决于输入提示内容（Prompt Content）的质量，因此如何针对图书推荐任务构建有效的提示工程（Prompt Engineering）[21] 是本小节的讨论核心。如图3所示，为用户图书评分偏好预估任务 Prompt 工程样例，通常包括四个核心部分：

(1) **身份注入提示**。该提示主要用以提示 LLM 当前所代表的角色类型，引导 LLM 按照特定的角色类型去作出不同的行为响应。例如，在某些特定任务中，出于安全或公平性的限制，如果不进行身份注入提示“假设你是一个 xxx 职业的专家”，而是直接要求 ChatGPT 进行回答例如：“请对《xxx》评分偏好进行评估”的任务要求，ChatGPT 通常会发出拒绝回答的响应。

(2) **任务描述提示**。如果说身份注入提示是为 ChatGPT 类模型设定可能的行为簇，而任务描述提示则是提示 LLM 当前所需要完成的具体任务背景、任务框架、以及可能的任务样例（Few-shot 场景）。通常情况下，基于任务样例（Few-shot 场景）的内容提示，相当于给模型增加了一定的学习样本，能够进一步增强模型对任务的拟合与理解效果，从而最终产出更好的预测结果。

身份注入
任务规则描述
假设你是一个专业的用户兴趣推荐专家，需要你对用户A的书籍偏好进行评分，评分范围在1-5分，1表示用户A不喜欢该书籍，5表示用户A非常喜欢该书籍。已知用户A自己对部分书籍评分结果：
由作者 Audrey 写的《The Time Traveler's Wife》，评分：4.00
由作者 Andy Weir 写的《The Martian》，评分：1.00
由作者 Neal Stephenson 写的《Seveneves》，评分：5.00
N-FewShot 内容提示
请对以下书籍进行评分，预测出用户对这些书籍的喜好。
由作者 Orson Scott Card 写的《Xenocide》
由作者 Christopher Paolini 写的《Eragon》
待打分书籍列表，同时对多篇进行打分
无需其他任何文字说明、解释，每行仅输出数字打分结果，每个评分保留两位小数
推荐输出标准化定义

图 3 用户图书评分偏好预估任务 Prompt 工程样例

（3）**任务边界提示**。上述提示工程主要从正向告知 ChatGPT 类模型需要做什么任务，而任务边界提示则主要是用以从负面限定模型：不要做什么。在本文任务中，如果仅使用了身份注入以及任务描述，模型通常会产出对应的评分以及大段的解释性语言，这会给后续应用带来困难。因此，还需要明确的限定并告诉模型，该任务的边界是什么，即：不需要任何文字解释，只需输出评分结果。此时，模型则会按照要求仅产出对应的评估分值。

（4）**输出格式提示**。在完成角色注入、任务描述以及边界提示之后，还需要最终告诉模型：所需要产出的数据格式。这一部分是为了方便将模型的产出结果与后续任务链路更好的结合。如，针对用户偏好评分任务，需要限定产出格式为：保留 2 位小数的数值。

3.4 模型交互与响应解析与校验

通过提示工程的构建，我们能够在一定程度上保障模型的输出符合预期的要求。但由于 ChatGPT 类模型本质是一种自然语言概率模型，同时，为了保障模型生成结果的多样性，模型在设计的过程中便加入了随机性因素[1]，这也可能使得模型对于相同的输入请求，产出不同的响应结果。因此，对于 ChatGPT 类模型的生成内容，我们还需要进一步构建“输出结果校验与任务重启”模块，对关键的产出数据格式与要求进行二次校验。

4 实验评估

4.1 数据集

Goodbook-10k 数据集。 GoodBook-10k[22] 数据集来源于 Goodreads[1] 书评网站（类似豆瓣读书），该网站是全球最大的在线读书社区。GoodBook-10k 数据集中包含有 1 万本热门图书与 598 万用户的图书评分数据，核心字段包括：图书评分、用户想读书籍标注、图书元数据、图书标签等。为了有效的对比模型在不同提示程度样本上的表现，我们将该数据集拆分为 1-Few-shot、10-Few-shot，以及 20-Few-shot 三种形式，即对应分别给模型提示该用户的 1 条、10 条、20 条评分记录，并以此构建训练集合、测试集合，以及 prompt 集合，要求模型对剩余记录进行预测，并给出用户的偏好排序。数据集的详细拆分逻辑如图4所示。

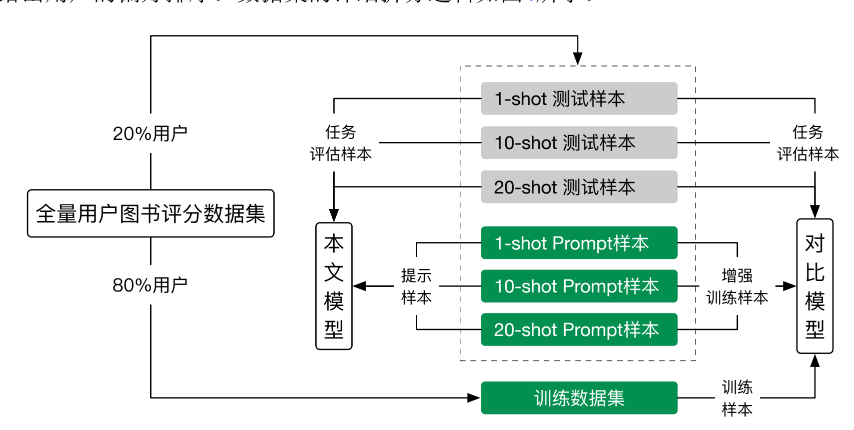


图 4 数据集评估方案拆分

4.2 对比模型

为了有效测试本文提出模型与现有推荐模型在用户图书评分偏好推荐场景下的表现差异，我们选取了个性化推荐场景的三个典型推荐算法模型：Matrix Factorization 模型（FunkSVD）[23]、KNN(means) 模型[24]，以及 SlopeOne[25] 模型。

[1]https://goodreads.com/

Matrix Factorization 模型（FunkSVD）[23] 是一种针对传统 SVD 模型在大规模数据场景下面临的计算效率和稀疏性难点所提出的改进模型。该模型能够将用户和图书的评分兴趣网络分解为用户矩阵和图书矩阵，即将用户和图书的关联特征都映射到一个 k 维度空间中，并基于映射的矩阵表征用户的兴趣偏好。

KNN(means) 模型[24] 则是通过考虑用户评分均值对于其偏好的影响来改进基础 KNN 模型的推荐策略。通过这种建模形式，能够保证预估得到的用户评分偏好会更加关注用户自身的评分分布，从而贴合现实应用场景。

SlopeOne[25] 模型则是一种非常经典且简洁的协同过滤推荐算法，该模型计算效率高，且易于对用户的潜在相似兴趣偏好进行建模。但由于模型建模十分依赖用户自身行为的丰富程度，当训练集中用户行为较少（如提示样本不够的情况下）时，其表现效果通常不佳。

4.3 评估指标

基于第3.2小节可以看到：用户图书评分偏好推荐问题既可以被看成是一个回归问题，也可以看成是一个排序问题。因此，为了验证基于模型在不同测试样本上的性能表现，我们将同时围绕回归以及排序模型的以下指标对模型效果进行评估：

指标一：平均绝对误差（MAE，Mean Absolute Error），考虑不同模型对于用户评分偏好预估结果的绝对偏差，关注真实值和预测值绝对误差的平均值。计算方式如下，

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \tilde{y}_i|$$

指标二：平均绝对百分比误差 (MAPE，Mean Absolute Percentage Error)，通过量纲缩放的方式，更多关注预估误差相对每个样本真实值的百分比偏差情况。计算方式如下，

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \tilde{y}_i| \,/\, |y_i|$$

指标三：均方根误差（RMSE，Root Mean Square Error），与 MAE 的关注点有所差异的是，RMSE 更加关注不同大小误差相对权重对模型带来的影响。其计算方式如下，

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2}$$

指标四：归一化折损累计增益（NDCG，Normalized Discounted Cumulative Gain）[26]，该指标主要用以观测在排序结果中相对位置的差异性好坏。在本文中，我们将分别考虑 NDCG@{5,10,15,20} 位置之前的表现结果。

4.4 结果分析与讨论

本小节将对 CUBR 模型以及对照模型在不同任务上的表现结果进行分析，核心回答两个问题，问题（1）：CUBR 模型能否在用户图书评分偏好推荐场景取得效果？与其他推荐模型相比效果怎么样？问题（2）：提示样本的增加，能否提高 CUBR 模型的推荐能力？与对比模型相比是否有明显变化？

表 1 用户评分偏好预估模型对比效果

建模方式	对比模型	评估指标						
		NDCG@5	NDCG@10	NDCG@15	NDCG@20	MAE	MAPE	RMSE
1-Few-shot	MF（FunkSVD）[23]	0.8764	0.8934	0.9186	0.9578	0.7657	0.2599	0.9566
	KNN（means）[24]	0.8427	0.8679	0.8997	0.9465	0.8655	0.2830	1.1536
	SlopeOne[25]	0.8298	0.8577	0.8919	0.9421	0.8438	0.2788	1.1177
	CUBR	0.8508	0.8740	0.9026	0.9496	1.0756	0.2977	1.3421
10-Few-shot	MF（FunkSVD）[23]	0.8753	0.8925	0.9180	0.9575	0.7333	0.2490	0.9179
	KNN（means）[24]	0.8802	0.8966	0.9210	0.9592	0.7081	0.2367	0.9124
	SlopeOne[25]	0.8635	0.8854	0.9124	0.9536	0.7410	0.2423	0.9650
	CUBR	0.8685	0.8839	0.9096	0.9541	0.9159	0.2634	1.1849
20-Few-shot	MF（FunkSVD）[23]	0.8759	0.8931	0.9184	0.9577	0.7108	0.2413	0.8928
	KNN（means）[24]	0.8824	0.8985	0.9223	0.9599	0.6769	0.2265	0.8792
	SlopeOne[25]	0.8718	0.8909	0.9166	0.9564	0.7000	0.2331	0.9051
	CUBR	0.8742	0.8896	0.9127	0.9565	0.8425	0.2485	1.1161

如表1 所示，为 CUBR 模型以及对照模型在 1-Few-shot,10-Few-shot, 和 20-Few-shot 三个子任务上的测试结果，其中灰色背景的数字为该子任务下的最优模型，下划线数字则对应次优模型。

首先，从整体上看：MF（FunkSVD）模型能够在不同子任务上都取得不错的效果，特别是在 1-Shot 的场景下取得了最优。其核心原因在于：基于 FunkSVD 的推荐策略是通过矩阵分解的方式，对用户-图书的评分交互矩阵

进行建模，这一建模方式的优化目标是让用户评分与矩阵乘积得到的评分残差尽可能的小，因此在待预测用户提供的参照评分信息有限的情况下（如 1-Few-shot）FunkSVD 在 RMSE 等指标上也能取得不错的效果。但随着单个用户有效提示样例（特征）的增多，以 KNN（means）为代表的聚类式模型则开始发挥出优势，在预测的过程中，KNN（means）会依赖于待预测用户的历史评分习惯建模来生成最终的预估结果，使得随着提示样例的增多，其预估的准确性也逐步增加。值得注意的是：CUBR 模型在单样本的场景下，基于 NDCG 指标上也取得了次优的推荐结果，说明 CUBR 在小样本推荐场景下的具有较好的泛化能力。但同时也看到：**在用户维度的个性化理解与建模上，直接应用通用 LLM 构建的 CUBR 模型相比经典推荐模型的预测效果还有一定差距。**

其次，具体到不同提示程度的子任务上看：FunkSVD 模型的 NDCG 指标对于提示样本的数量并不敏感，在 1-few-shot、10-few-shot 和 20-few-shot 的子任务上的其 NDCG 结果表现基本一致，但其余对照模型，如 KNN（means）、SlopeOne 和 CUBR 随着待预测样本提示数量的增加，其效果指标都发生了较大的变化。核心原因在于后续模型在结果预测的过程中，会对待预测样本用户的历史打分进行参照。例如，CUBR 模型在打分的过程中，会参照提供的用户针对不同书籍的历史打分结果，以此建模用户的兴趣偏好的相关背景知识，并在新的待预测的样本中参照该知识信息进行综合评估。如图5所示，为 CUBR 外显的打分决策过程。通过该样例可以看到，基于恰当的样例提示，LLM 模型通常能够学习到对应的语境知识并在预测场景进行应用，该能力又称为语境学习能力（In-context Learning）[1]，也是大语言模型重要的基础能力之一。通过增加提示样本参照之后，在 20-Few-shot 子任务上，CUBR 的 MAE、MAPE 和 RMSE 指标都有显著改善，误差分别降低：21.67%、16.53% 和 16.84%。

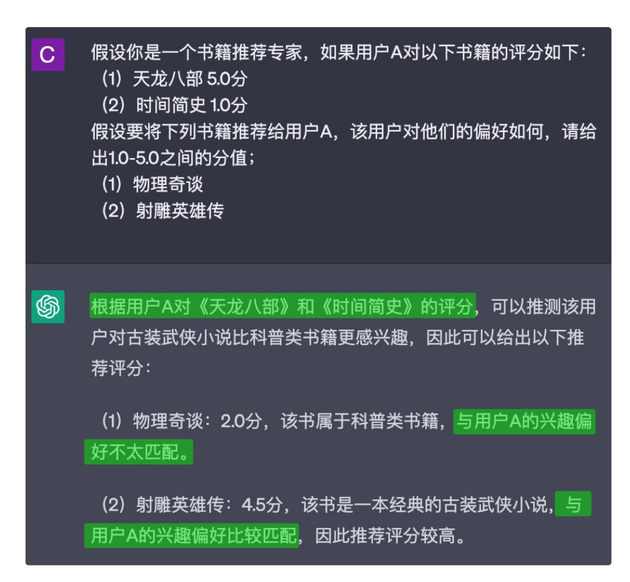


图 5 基于小样本提示的用户评分偏好预估样例

最后，从不同类型的对照指标上看：在参照样本较少的情况下（如 1-Few-shot），CUBR 模型虽然在排序能力（NDCG）上的效果较好，但在分值预估的误差上度量上（MAE/MAPE/RMSE），相比对照模型均无优势。此外，对于单个用户提示样本的过多增加（如从 10-Few-shot 增加到 20-Few-shot），从 NDCG 指标上看，并不会引起包括 CUBR 在内的模型有类似从 1-Few-shot 到 10-Few-shot 的效果增长，但对于误差类指标而言，其涨幅依然可观。因此，如果应用场景更加看中每个待推荐用户的绝对分值偏好，可以通过增加待预测用户提示样本的方式提升效果。如果场景仅关注相对排序能力，则基于少数样本进行建模即可满足要求，可以进一步节省推理资源。

5 总结与未来工作

本文提出了一种基于 ChatGPT 类大语言技术的用户图书评分偏好预测模型（CUBR），该模型首次将 LLM 技术引入图情领域的经典任务并落地实践。在用户图书评分偏好预测任务中，我们分别在 1-Few-shot、10-Few-shot 和 20-Few-shot 三个不同样本提示程度的子任务上进行了测试。实验结果表明：CUBR 在提示样本较少且未进行任何微调的情况下，能够取得不错的推荐效果，且通过增加提示样本的数量后，其预测结果提升明显。未来，我们将持续围绕以下方面进行研究：

研究点一：融合多源数据的 Prompt 构建研究。在当前的探索中，我们仅用到了用户的评分交互数据用以构建 Prompt 集合。在后续研究中，如何融合来自多源的用户属性与特征，甚至跨模态数据，并在 LLM 中进行表达，由此构建统一的推荐系统，值得深入的探索。基本上思路是：如何更好的把来自多源异构的特征数据，一致的表达为 LLM 可以理解的自然语言编码，从而充分利用各类语境信息，进一步提高模型的预测效果。

研究点二：面向任务指令微调的建模研究。当前 CUBR 的建模思路是直接应用已训练完成的 LLM 进行推荐与应用，这种形式通常考验的是 LLM 的泛化能力，但已有研究表明：通过针对性的构建指令训练集，能够进一步增加 LLM 在特定任务上的表现效果[1]。因此，在未来的研究中，如何考虑基于图情领域的专有任务以及特殊业务场景，构建高效的微调指令集，并将预测过程与训练过程联合起来，最终提高 LLM 模型在推荐系统中的表现，也非常值得探讨。

参考文献

[1] ZHAO W X, ZHOU K, LI J, et al. A survey of large language models[A]. 2023: 1-58.

[2] OPENAI. Introducing chatgpt[R/OL]. https://openai.com/blog/chatgpt.

[3] DREIBELBIS E. Chatgpt passes google coding interview for level 3 engineer with $183k salary[EB/OL]. http://985.so/mny2k.

[4] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.

[5] TOUVRON H, LAVRIL T, IZACARD G, et al. Llama: Open and efficient foundation language models[A]. 2023.

[6] ZENG A, LIU X, DU Z, et al. GLM-130b: An open bilingual pre-trained model[C/OL]//The Eleventh International Conference on Learning Representations (ICLR). 2023. https://openreview.net/forum?id=-Aw0rrrPUF.

[7] GEORGE A S, GEORGE A H. A review of chatgpt ai's impact on several business sectors[J]. Partners Universal International Innovation Journal, 2023, 1(1): 9-23.

[8] LU Q, QIU B, DING L, et al. Error analysis prompting enables human-like translation evaluation in large language models: A case study on chatgpt[A]. 2023.

[9] SHAFEEG A, SHAZHAEV I, MIHAYLOV D, et al. Voice assistant integrated with chat gpt[J]. Indonesian Journal of Computer Science, 2023, 12(1).

[10] 游俊哲. ChatGPT 类生成式人工智能在科研场景中的应用风险与控制措施[J]. 情报理论与实践, 2023: 01-11.

[11] 蔡士林，杨磊. ChatGPT 智能机器人应用的风险与协同治理研究[J]. 情报理论与实践, 2023: 01-11.

[12] 张慧，佟彤，叶鹰. AI 2.0 时代智慧图书馆的 GPT 技术驱动创新[J]. 图书馆杂志, 2023: 01-07.

[13] 郭亚军，郭一若，李帅，冯思倩. ChatGPT 赋能图书馆智慧服务：特征、场景与路径[J]. 图书馆建设, 2023: 01-16.

[14] 周文欢. ChatGPT 在档案领域应用和意义[J]. 中国档案, 2023, 593(03): 62-63.

[15] 汪波，牛朝文. 从 ChatGPT 到 GovGPT：生成式人工智能驱动的政务服务生态系统构建[J]. 电子政务, 2023: 01-14.

[16] 余以胜，韦锐，刘鑫艳. 可解释的实时图书信息推荐模型研究[J]. 情报学报, 2019, 38(2): 209-216.

[17] 杨辰，刘婷婷，刘雷，牛奔，孙见山. 融合语义和社交特征的电子文献资源推荐方法研究[J]. 情报学报, 2019, 38(6): 632-640.

[18] WANG S, HU L, WANG Y, et al. Sequential recommender systems: Challenges, progress and prospects[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19. International Joint Conferences on Artificial Intelligence Organization, 2019: 6332-6338.

[19] 王代琳，刘丽娜，刘美玲，刘亚秋. 基于图书目录注意力机制的读者偏好分析与推荐模型研究[J]. 数据分析与知识发现, 2022, 6(9): 138-152.

[20] 陈帜，张文德，刘田. 基于图卷积神经网络的图书推荐方法研究[J]. 情报探索, 2022, 300(10): 1-5.

[21] SARAVIA E. Prompt Engineering Guide[J]. https://www.promptingguide.ai, 2022.

[22] ZAJAC Z. Goodbooks-10k: a new dataset for book recommendations[J/OL]. FastML, 2017. http://fastml.com/goodbooks-10k.

[23] MNIH A, SALAKHUTDINOV R R. Probabilistic matrix factorization[J]. Advances in neural information processing systems （NPIS）, 2007, 20.

[24] KOREN Y. Factor in the neighbors: Scalable and accurate collaborative filtering[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2010, 4(1): 1-24.

[25] LEMIRE D, MACLACHLAN A. Slope one predictors for online rating-based collaborative filtering[C]//Proceedings of the 2005 SIAM International Conference on Data Mining. SIAM, 2005: 471-475.

[26] WANG Y, WANG L, LI Y, et al. A theoretical analysis of ndcg type ranking measures[C]//Conference on learning theory. PMLR, 2013: 25-54.

作者简介（一）：陈燕方，女，1992，馆员，博士。研究方向：用户行为、健康信息、数据挖掘。

作者简介（二）：李志宇（通信作者：zhiyulee@icloud.com），男，1991，算法专家，博士。研究方向：机器学习、网络表征、自然语言处理。

作者贡献声明：陈燕方，负责论文起草，研究框架制定，主体内容撰写。李志宇，负责实验数据收集、模型实现，实验讨论与分析。